An Improved Graph-Entropy Bound for Perfect Hashing *

Erdal Arikan
Electrical Engineering Department, Bilkent University, 06533 Ankara, Turkey

Abstract

We give an improved graph-entropy bound on the size of families of perfect hash functions. Examples are given illustrating that the new bound improves previous bounds in several instances.

Perfect hashing is a method of information storage and retrieval [1]. It is also equivalent to certain zero-error list-coding problems of information theory [4]. Following [3], call a set of sequences of length t over a b-letter alphabet k-*separated* if for every k-tuple of sequences there exists a coordinate in which they all differ. Let $N(t,b,k)$ denote the largest possible size for such a set of sequences. Perfect-hashing is the problem of finding such maximal sets. Here, we give an upper bound on the asymptotic quantity (the capacity)

$$C_{b,k} := \limsup_{t\to\infty} \frac{1}{t} \log N(t,b,k)$$

(Logarithms are to base 2.) This bound extends earlier results given in [5] and is based on a refinement of the graph-entropy bound [2]-[4].

Körner and Marton [3] show that

$$C_{b,k} \le \min_{0\le j\le k-2} \frac{b^{\underline{j+1}}}{b^{j+1}} \log \frac{b-j}{k-j-1} \tag{1}$$

where $b^{\underline{i}} = b(b-1)\cdots(b-i+1)$. We give here the bound

$$C_{b,k} \le \sup\{x : x \le \alpha_j(x),\ j = 2,\ldots,k-2\} \tag{2}$$

where

$$\alpha_j(x) = \frac{b-j}{k-1} 2^{-x} \left(1 - \frac{x}{\log b}\right) \frac{b^{\underline{j}}}{b^j} \log \frac{b-j}{k-1-j}$$

for $j = 2,\ldots,b-k$ and

$$\alpha_j(x) = \left(1 - \frac{j}{b-k+1}(1-2^{-x})\right)\left(1 - \frac{x}{\log b}\right) \frac{b^{\underline{j}}}{b^j} \log \frac{b-j}{k-1-j}$$

for $j = b-k+1,\ldots,k-2$.

The following table lists the values of the new bound (2) and the Körner-Marton bound (1). The integers in parentheses indicate the values of j which optimize the corresponding bounds. The table demonstrates that the new bound improves the earlier graph-entropy bound in many instances. To our knowledge, the values in the table constitute the best available bounds on $C_{b,k}$.

b	k	New Bound	KM Bound
4	4	0.3511 (2)	0.3750 (2)
5	4	0.6114 (2)	0.7370 (0)
5	5	0.2359 (3)	0.1920 (3)
6	4	0.8390 (2)	1.0000 (0)
6	5	0.4414 (3)	0.4402 (3)
6	6	0.1548 (4)	0.0925 (4)
7	4	1.029 (2)	1.2223 (0)
7	5	0.6204 (3)	0.6997 (3)
7	6	0.3055 (4)	0.2376 (4)
7	7	0.0974 (5)	0.0428 (5)
100	6	3.6184 (2)	4.3219 (0)
100	10	2.830 (2)	3.3219 (0)

References

[1] M. Fredman and J. Komlós, 'On the size of separating systems and perfect hash functions,' *SIAM J. Algebraic and Discrete Methods*, vol. 5, no. 1, pp. 61-68, 1984.

[2] J. Körner, 'Fredman-Komlós bounds and information theory,' *SIAM J. Algebraic and Discrete Methods*, vol. 7, no. 4, pp. 560-570, Oct. 1986.

[3] J. Körner and K. Marton, 'New bounds for perfect hashing via information theory,' *Europ. J. Combinatorics*, vol. 9, pp. 523-530, 1988.

[4] J. Körner and K. Marton, 'On the capacity of uniform hypergraphs,' *IEEE Trans. Inform. Theory*, vol. IT-36, No.1, pp. 153-156, Jan. 1990.

[5] E. Arikan, 'An upper bound on the zero-error list-coding capacity,' *Proceedings of 1993 IEEE Int. Symp. Inform. Theory*, p. 152, San Antonio, USA, Jan 17-22, 1993.

*This research was supported by TÜBİTAK under project TBAG 1053.